IFWO

```
                 RAW SEQUENCE LISTING                   DATE: 08/30/2004
                 PATENT APPLICATION:  US/10/681,207     TIME: 12:56:37

                 Input Set : N:\Crf3\RULE60\10681207.raw
                 Output Set: N:\CRF4\08302004\J681207.raw

     1 <110> APPLICANT: Merck Patent GmbH
     2 <120> TITLE OF INVENTION: Glucose Dehydrogenase Fusion Proteins and their Utilization
in Expression
     3       Systems
     4 <130> FILE REFERENCE: Merck 2289 / P9906920
     5 <140> CURRENT APPLICATION NUMBER: US/10/681,207
     6 <141> CURRENT FILING DATE: 2003-10-09
     7 <150> PRIOR APPLICATION NUMBER: US/09/913,494
     8 <151> PRIOR FILING DATE: 2001-08-16
     9 <150> PRIOR APPLICATION NUMBER: DE 19906920
    10 <151> PRIOR FILING DATE: 1999-02-19
    11 <160> NUMBER OF SEQ ID NOS: 17
    12 <170> SOFTWARE: PatentIn version 3.1
    14 <210> SEQ ID NO: 1
    15 <211> LENGTH: 3992
    16 <212> TYPE: DNA
    17 <213> ORGANISM: Bacillus megaterium
    18 <220> FEATURE:
    19 <221> NAME/KEY: CDS
    20 <222> LOCATION: (186)..(968)
    21 <223> OTHER INFORMATION: Glucose Dehydrogenase from Bacillus megaterium
    22 <220> FEATURE:
    23 <221> NAME/KEY: gene
    24 <222> LOCATION: (1)..(3992)
    25 <223> OTHER INFORMATION: plasmid PAW2
    26 <220> FEATURE:
    27 <221> NAME/KEY: CDS
    28 <222> LOCATION: (978)..(1010)
    29 <223> OTHER INFORMATION: poly-histidin tag
    30 <400> SEQUENCE: 1
    31       ccatcgaatg gccagatgat taattcctaa tttttgttga cactctatca ttgatagagt      60
    32       tattttacca ctccctatca gtgatagaga aaagtgaaat gaatagttcg acaaaaatct     120
    33       agataacgag ggcaatcgat gaattcgagc tcggtacccg gggatccctc gaggtcgacc     180
    34       tgcag atg tat aca gat tta aaa gat aaa gta gtt gta att aca ggt gga     230
    35             Met Tyr Thr Asp Leu Lys Asp Lys Val Val Val Ile Thr Gly Gly
    36             1               5                   10                  15
    37       tca aca ggt tta gga cgc gca atg gct gtt cgt ttc ggt caa gaa gaa     278
    38       Ser Thr Gly Leu Gly Arg Ala Met Ala Val Arg Phe Gly Gln Glu Glu
    39                       20                  25                  30
    40       gca aaa gtt gtt att aac tat tac aac aat gaa gaa gaa gct cta gat     326
    41       Ala Lys Val Val Ile Asn Tyr Tyr Asn Asn Glu Glu Glu Ala Leu Asp
    42                   35                  40                  45
    43       gcg aaa aaa gaa gta gaa gaa gca ggc gga caa gca atc atc gtt caa     374
    44       Ala Lys Lys Glu Val Glu Glu Ala Gly Gly Gln Ala Ile Ile Val Gln
```

ENTERED

```
        50                  55                  60
ggc gat gta aca aaa gaa gaa gac gtt gta aat ctt gtt caa aca gct      422
Gly Asp Val Thr Lys Glu Glu Asp Val Val Asn Leu Val Gln Thr Ala
    65                  70                  75
att aaa gaa ttt ggt aca tta gac gta atg att aac aac gct ggt gtt      470
Ile Lys Glu Phe Gly Thr Leu Asp Val Met Ile Asn Asn Ala Gly Val
80                  85                  90                  95
gaa aac cca gtt cct tct cat gag cta tct cta gat aac tgg aac aaa      518
Glu Asn Pro Val Pro Ser His Glu Leu Ser Leu Asp Asn Trp Asn Lys
                100                 105                 110
gtt att gat aca aac tta aca ggt gca ttc tta gga agc cgt gaa gca      566
Val Ile Asp Thr Asn Leu Thr Gly Ala Phe Leu Gly Ser Arg Glu Ala
            115                 120                 125
att aaa tac ttc gtt gaa aac gac att aaa gga aat gtt atc aac atg      614
Ile Lys Tyr Phe Val Glu Asn Asp Ile Lys Gly Asn Val Ile Asn Met
        130                 135                 140
tct agc gtt cac gaa atg att cct tgg cca tta ttt gtt cac tac gca      662
Ser Ser Val His Glu Met Ile Pro Trp Pro Leu Phe Val His Tyr Ala
    145                 150                 155
gca agt aaa ggc ggt atg aaa cta atg acg gaa aca ttg gct ctt gaa      710
Ala Ser Lys Gly Gly Met Lys Leu Met Thr Glu Thr Leu Ala Leu Glu
160                 165                 170                 175
tat gcg cca aaa ggt atc cgc gta aat aat att gga cca ggt gcg atg      758
Tyr Ala Pro Lys Gly Ile Arg Val Asn Asn Ile Gly Pro Gly Ala Met
                180                 185                 190
aac aca cca att aac gca gag aaa ttt gca gat cca gaa caa cgt gca      806
Asn Thr Pro Ile Asn Ala Glu Lys Phe Ala Asp Pro Glu Gln Arg Ala
            195                 200                 205
gac gta gaa agc atg att cca atg ggt tac atc ggt aaa cca gaa gaa      854
Asp Val Glu Ser Met Ile Pro Met Gly Tyr Ile Gly Lys Pro Glu Glu
        210                 215                 220
gta gca gca gtt gca gca ttc tta gct tca tca caa gca agc tat gta      902
Val Ala Ala Val Ala Ala Phe Leu Ala Ser Ser Gln Ala Ser Tyr Val
    225                 230                 235
aca ggt att aca tta ttt gca gat ggc ggt atg acg aaa tac cct tct      950
Thr Gly Ile Thr Leu Phe Ala Asp Gly Gly Met Thr Lys Tyr Pro Ser
240                 245                 250                 255
ttc caa gca gga aga ggc taatagagc gct atg aga gga tcg cat cac cat   1001
Phe Gln Ala Gly Arg Gly           Ala Met Arg Gly Ser His His His
                260                           265
cac cat cac taatagaagc ttgacctgtg aagtgaaaaa tggcgcacat             1050
His His His
270
tgtgcgacat tttttttgtc tgccgtttac cgctactgcg tcacggatct ccacgcgccc   1110
tgtagcggcg cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt   1170
gccagcgccc tagcgcccgc tcctttcgct ttcttccctt cctttctcgc cacgttcgcc   1230
ggctttcccc gtcaagctct aaatcggggg ctccctttag ggttccgatt tagtgcttta   1290
cggcacctcg accccaaaaa acttgattag ggtgatggtt cacgtagtgg gccatcgccc   1350
tgatagacgg tttttcgccc tttgacgttg gagtccacgt tctttaatag tggactcttg   1410
```

```
94        ttccaaactg gaacaacact caaccctatc tcggtctatt cttttgattt ataagggatt   1470
95        ttgccgattt cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat   1530
96        tttaacaaaa tattaacgct tacaatttca ggtggcactt ttcggggaaa tgtgcgcgga   1590
97        acccctattt gtttattttt ctaaatacat tcaaatatgt atccgctcat gagacaataa   1650
98        ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca acatttccgt   1710
99        gtcgccctta ttcccttttt tgcggcattt tgccttcctg tttttgctca cccagaaacg   1770
100        ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta catcgaactg   1830
101        gatctcaaca gcggtaagat ccttgagagt tttcgccccg aagaacgttt tccaatgatg   1890
102        agcactttta aagttctgct atgtggcgcg gtattatccc gtattgacgc cgggcaagag   1950
103        caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc accagtcaca   2010
104        gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgctgc cataaccatg   2070
105        agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa ggagctaacc   2130
106        gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga accggagctg   2190
107        aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg   2250
108        ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca attgatagac   2310
109        tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc ggctggctgg   2370
110        tttattgctg ataaatctgg agccggtgag cgtggctctc gcggtatcat tgcagcactg   2430
111        gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag tcaggcaact   2490
112        atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa gcattggtag   2550
113        gaattaatga tgtctcgttt agataaaagt aaagtgatta acagcgcatt agagctgctt   2610
114        aatgaggtcg gaatcgaagg tttaacaacc cgtaaactcg cccagaagct aggtgtagag   2670
115        cagcctacat tgtattggca tgtaaaaaat aagcgggctt tgctcgacgc cttagccatt   2730
116        gagatgttag ataggcacca tactcacttt tgccctttag aaggggaaag ctggcaagat   2790
117        tttttacgta ataacgctaa aagttttaga tgtgctttac taagtcatcg cgatggagca   2850
118        aaagtacatt taggtacacg gcctacagaa aaacagtatg aaactctcga aaatcaatta   2910
119        gcctttttat gccaacaagg tttttcacta gagaatgcat tatatgcact cagcgcagtg   2970
120        gggcatttta ctttaggttg cgtattggaa gatcaagagc atcaagtcgc taaagaagaa   3030
121        agggaaacac ctactactga tagtatgccg ccattattac gacaagctat cgaattattt   3090
122        gatcaccaag gtgcagagcc agccttctta ttcggccttg aattgatcat atgcggatta   3150
123        gaaaaacaac ttaaatgtga aagtgggtct taaaagcagc ataacctttt tccgtgatgg   3210
124        taacttcact agtttaaaag gatctaggtg aagatccttt ttgataatct catgaccaaa   3270
125        atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa gatcaaagga   3330
126        tcttcttgag atcctttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg   3390
127        ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctctttttcc gaaggtaact   3450
128        ggcttcagca gagcgcagat accaaatact gtccttctag tgtagccgta gttaggccac   3510
129        cacttcaaga actctgtagc accgcctaca tacctcgctc tgctaatcct gttaccagtg   3570
130        gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg   3630
131        gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga   3690
132        acgacctaca ccgaactgag atacctacag cgtgagctat gagaaagcgc cacgcttccc   3750
133        gaagggagaa aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg   3810
134        agggagcttc cagggggaaa cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc   3870
135        tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc   3930
136        agcaacgcgg cctttttacg gttcctggcc ttttgctggc cttttgctca catgacccga   3990
137        ca                                                                    3992
139 <210> SEQ ID NO: 2
140 <211> LENGTH: 261
141 <212> TYPE: PRT
142 <213> ORGANISM: Bacillus megaterium
143 <400> SEQUENCE: 2
```

```
      Met Tyr Thr Asp Leu Lys Asp Lys Val Val Val Ile Thr Gly Gly Ser
      1               5                   10                  15
      Thr Gly Leu Gly Arg Ala Met Ala Val Arg Phe Gly Gln Glu Glu Ala
                  20                  25                  30
      Lys Val Val Ile Asn Tyr Tyr Asn Asn Glu Glu Glu Ala Leu Asp Ala
              35                  40                  45
      Lys Lys Glu Val Glu Glu Ala Gly Gly Gln Ala Ile Ile Val Gln Gly
          50                  55                  60
      Asp Val Thr Lys Glu Glu Asp Val Val Asn Leu Val Gln Thr Ala Ile
      65                  70                  75                  80
      Lys Glu Phe Gly Thr Leu Asp Val Met Ile Asn Asn Ala Gly Val Glu
                      85                  90                  95
      Asn Pro Val Pro Ser His Glu Leu Ser Leu Asp Asn Trp Asn Lys Val
                  100                 105                 110
      Ile Asp Thr Asn Leu Thr Gly Ala Phe Leu Gly Ser Arg Glu Ala Ile
              115                 120                 125
      Lys Tyr Phe Val Glu Asn Asp Ile Lys Gly Asn Val Ile Asn Met Ser
          130                 135                 140
      Ser Val His Glu Met Ile Pro Trp Pro Leu Phe Val His Tyr Ala Ala
      145                 150                 155                 160
      Ser Lys Gly Gly Met Lys Leu Met Thr Glu Thr Leu Ala Leu Glu Tyr
                      165                 170                 175
      Ala Pro Lys Gly Ile Arg Val Asn Asn Ile Gly Pro Gly Ala Met Asn
                  180                 185                 190
      Thr Pro Ile Asn Ala Glu Lys Phe Ala Asp Pro Glu Gln Arg Ala Asp
              195                 200                 205
      Val Glu Ser Met Ile Pro Met Gly Tyr Ile Gly Lys Pro Glu Glu Val
          210                 215                 220
      Ala Ala Val Ala Ala Phe Leu Ala Ser Ser Gln Ala Ser Tyr Val Thr
      225                 230                 235                 240
      Gly Ile Thr Leu Phe Ala Asp Gly Gly Met Thr Lys Tyr Pro Ser Phe
                      245                 250                 255
      Gln Ala Gly Arg Gly
                  260
<210> SEQ ID NO: 3
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Bacillus megaterium
<400> SEQUENCE: 3
      Ala Met Arg Gly Ser His His His His His His
      1               5                   10
<210> SEQ ID NO: 4
<211> LENGTH: 4193
<212> TYPE: DNA
<213> ORGANISM: Bacillus megaterium / Heamenteria ghilianii fusion
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (141)..(344)
<223> OTHER INFORMATION: Tridegin
```

```
<220> FEATURE:
<221> NAME/KEY: gene
<222> LOCATION: (1)..(4193)
<223> OTHER INFORMATION: plasmid pAW4
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (387)..(1169)
<223> OTHER INFORMATION: Glucose Dehydrogenase
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1179)..(1211)
<223> OTHER INFORMATION: poly histidine tag
<400> SEQUENCE: 4
      ccatcgaatg gccagatgat taattcctaa tttttgttga cactctatca ttgatagagt     60
      tattttacca ctccctatca gtgatagaga aaagtgaaat gaatagttcg acaaaaatct    120
      agataacgag ggcaatcgat atg aaa cta ttg cct tgc aaa gaa tgg cat caa    173
                            Met Lys Leu Leu Pro Cys Lys Glu Trp His Gln
                            1               5                   10
      ggt att cct aac cct agg tgc tgg tgt ggg gct gat cta gaa tgc gca      221
      Gly Ile Pro Asn Pro Arg Cys Trp Cys Gly Ala Asp Leu Glu Cys Ala
                  15                  20                  25
      caa gac caa tac tgt gcc ttc ata cct caa tgt aga cca aga tca gaa      269
      Gln Asp Gln Tyr Cys Ala Phe Ile Pro Gln Cys Arg Pro Arg Ser Glu
              30                  35                  40
      ctg att aaa cct atg gat gat ata tac caa aga cca gtc gag ttt cca      317
      Leu Ile Lys Pro Met Asp Asp Ile Tyr Gln Arg Pro Val Glu Phe Pro
          45                  50                  55
      aac ctt cca tta aaa cct agg gag gaa agcgctatga gaggatcgca            364
      Asn Leu Pro Leu Lys Pro Arg Glu Glu
      60                  65
      tcaccatcac catcacctgc ag atg tat aca gat tta aaa gat aaa gta gtt     416
                               Met Tyr Thr Asp Leu Lys Asp Lys Val Val
                                   70                  75
      gta att aca ggt gga tca aca ggt tta gga cgc gca atg gct gtt cgt      464
      Val Ile Thr Gly Gly Ser Thr Gly Leu Gly Arg Ala Met Ala Val Arg
          80                  85                  90
      ttc ggt caa gaa gaa gca aaa gtt gtt att aac tat tac aac aat gaa      512
      Phe Gly Gln Glu Glu Ala Lys Val Val Ile Asn Tyr Tyr Asn Asn Glu
      95                  100                 105                 110
      gaa gaa gct cta gat gcg aaa aaa gaa gta gaa gaa gca ggc gga caa      560
      Glu Glu Ala Leu Asp Ala Lys Lys Glu Val Glu Glu Ala Gly Gly Gln
                      115                 120                 125
      gca atc atc gtt caa ggc gat gta aca aaa gaa gaa gac gtt gta aat      608
      Ala Ile Ile Val Gln Gly Asp Val Thr Lys Glu Glu Asp Val Val Asn
                  130                 135                 140
      ctt gtt caa aca gct att aaa gaa ttt ggt aca tta gac gta atg att      656
      Leu Val Gln Thr Ala Ile Lys Glu Phe Gly Thr Leu Asp Val Met Ile
              145                 150                 155
      aac aac gct ggt gtt gaa aac cca gtt cct tct cat gag cta tct cta      704
```

```
RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 08/30/2004
PATENT APPLICATION:  US/10/681,207      TIME: 12:56:38

Input Set : N:\Crf3\RULE60\10681207.raw
Output Set: N:\CRF4\08302004\J681207.raw
```

Invalid Line Length:
The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:1; Line(s) 2

```
VERIFICATION SUMMARY                          DATE: 08/30/2004
PATENT APPLICATION:  US/10/681,207            TIME: 12:56:38

Input Set : N:\Crf3\RULE60\10681207.raw
Output Set: N:\CRF4\08302004\J681207.raw

L:452 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:12,Line#:0
```